CUSTOMER BEHAVIOR PREDICTION

A PROJECT REPORT

Submitted by

PARTH RATHORE

UNI ROLL NO: 202401100300168

in partial fulfillment for the award of the degree

of

Bachelor of Technology – CSE Artificial Intelligence

Introduction

In today's highly competitive and dynamic market landscape, understanding and anticipating customer behavior has become a crucial strategy for businesses aiming to stay ahead. Customer behavior prediction involves analyzing past purchasing patterns, preferences, and interactions to forecast future actions such as buying decisions, product preferences, and customer loyalty.

This predictive capability enables companies to personalize marketing campaigns, optimize inventory management, enhance customer satisfaction, and ultimately drive revenue growth. With the increasing availability of large datasets and advanced machine learning tools, businesses can now uncover deep insights into customer motivations and trends, allowing for more accurate forecasting and strategic decision-making.

The ability to predict customer behavior not only helps in reducing churn and increasing retention but also enables companies to deliver products and services that align more closely with customer expectations. This report explores the methodologies involved in predicting customer behavior, highlighting the tools, techniques, and data-driven models used in this field.

Methodology

The methodology for customer behavior prediction typically involves several key steps, each designed to transform raw customer data into actionable insights. The following outlines the standard approach:

1. Data Collection

The first step involves gathering comprehensive customer data from various sources such as transaction records, website activity logs, social media interactions, customer surveys, and CRM systems. The data collected may include demographic details, purchase history, browsing patterns, feedback, and behavioral cues.

2. Data Preprocessing

Once collected, the raw data often requires cleaning and preprocessing. This step includes handling missing values, removing duplicates, normalizing data, and transforming categorical variables into numerical formats. Proper preprocessing ensures that the dataset is reliable and suitable for analysis.

3. Feature Selection and Engineering

Feature selection involves identifying the most relevant variables that influence customer behavior. Feature engineering, on the other hand, creates new variables from existing data to improve model accuracy. For example, calculating average purchase frequency or customer lifetime value (CLV) can provide more predictive power.

4. Model Selection

Several statistical and machine learning models can be applied to predict customer behavior. Common models include:

- **Logistic Regression** for classification problems like churn prediction.
- **Decision Trees and Random Forests** for understanding decision pathways.
- **Support Vector Machines (SVM)** for pattern recognition.
- **Neural Networks** for complex, non-linear relationships.
- **Clustering Algorithms** for customer segmentation.

The choice of model depends on the type of prediction (classification, regression, segmentation) and the nature of the dataset.

5. Model Training and Validation

The dataset is divided into training and testing subsets. The training data is used to fit the model, while the testing data validates its predictive accuracy. Cross-validation techniques are often employed to minimize overfitting and ensure that the model generalizes well to unseen data.

6. Prediction and Interpretation

Once validated, the model can predict customer behavior on new data. Interpretation of these predictions helps businesses make informed decisions — for instance, identifying high-value customers, anticipating churn, or suggesting personalized product recommendations.

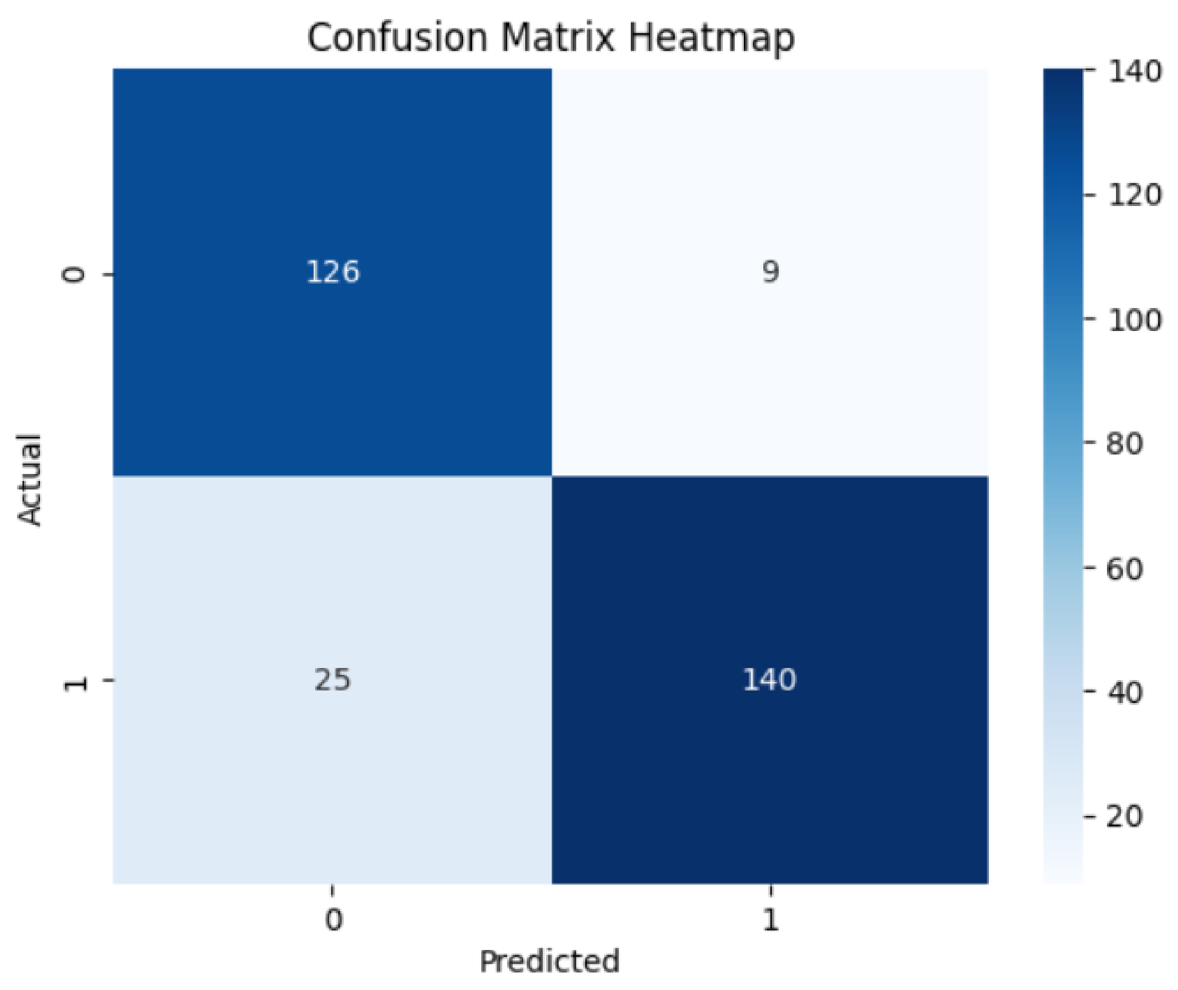
7. Continuous Improvement

As customer preferences and market conditions evolve, it’s important to retrain and refine prediction models with new data to maintain accuracy and relevance.

CODE

```
import pandas as pd

import numpy as np

from sklearn.model_selection import train_test_split

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import confusion_matrix, accuracy_score, precision_score, recall_score

import seaborn as sns

import matplotlib.pyplot as plt

# Example: loading dataset

# df = pd.read_csv('customer_data.csv')

# X = df.drop('customer_type', axis=1)

# y = df['customer_type']

# For illustration — generate dummy data

from sklearn.datasets import make_classification

X, y = make_classification(n_samples=1000, n_features=10, n_classes=2, random_state=42)

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Train model

model = RandomForestClassifier(random_state=42)

model.fit(X_train, y_train)

# Predictions

y_pred = model.predict(X_test)

# Confusion Matrix

cm = confusion_matrix(y_test, y_pred)

sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')

plt.xlabel('Predicted')

plt.ylabel('Actual')

plt.title('Confusion Matrix Heatmap')

plt.show()

# Metrics

accuracy = accuracy_score(y_test, y_pred)

precision = precision_score(y_test, y_pred)

recall = recall_score(y_test, y_pred)


print(f"Accuracy: {accuracy:.2f}")

print(f"Precision: {precision:.2f}")

print(f"Recall: {recall:.2f}")
```

OUTPUT




```
Accuracy: 0.89
Precision: 0.94
Recall: 0.85
```

REFERENCES

https://www.kaggle.com/code/gcdatkin/customer-behavior-prediction